Survival Analysis for Small Business: Evidence from Yelp

Ao Huang, Xiru Pan

**1 Introduction**

Accurately forecasting bankruptcy or business failure has long been a challenging issue for both researchers and practitioners, spanning several decades and across various fields like finance and accounting. Various methodologies, including ratio analysis (Beaver, 1966), Altman Z-scores (Altman, 1967), and Bayesian models (Sarkar and Sriram, 2001), have been employed to tackle this problem, all of which primarily depend on financial metrics such as stock prices, working capital, and levels of debt. These approaches are particularly valuable for medium to large enterprises, especially those that are publicly listed. However, they fall short when applied to small businesses like restaurants, which aren't publicly traded and lack readily available financial data for analysis. Past studies have often overlooked small businesses due to the challenge of accessing their financial information, compounded by the issue of such data being outdated. Traditional economic predictions usually rely on statistics from governmental bodies like the Census Bureau or the Department of Labor, but these statistics suffer from significant publication delays and a level of aggregation that may not be applicable to the needs of small businesses.

As information technology advances, particularly with the expansion of online review platforms, a vast amount of business-related data can now be gathered via the Internet, supporting the data-driven decision making. These data are particularly valuable since the online review platform can significantly influence people buying decision. The information on reviews platforms, often shared by previous customers, help potential buyers feel more confident about purchasing the products or services they are interested in (Duan et al. 2008). Numerous past studies have thoroughly examined the beneficial impact of online reviews on the performance of small businesses. For instance, the positive influence of online reviews on aspects of hotel performance such as sales, revenue per available room, and overall financial success has been well established in the literature (Anagnostopoulou et al., 2020; Nieto et al., 2014; Phillips et al., 2017). There's a positive relationship between the quantity of online reviews and customer satisfaction, as well as the reputation and financial outcomes of hospitality businesses (Nieto et al., 2014). Furthermore, review ratings are closely linked to key metrics like hotel occupancy rates, revenue per available room, sales figures, and the value of booking transactions (Torres et al., 2015; Viglia et al., 2016). However, despite the extensive examination of online reviews from various perspectives, there has been limited investigation into how these reviews impact the survival of businesses. Moreover, few studies consider additional business attributes such as the availability of amenities, parking options, or appointment requirements, which are crucial pieces of information available on online review platforms that customers often consider before deciding to visit a business.

In this research, we concentrate on predicting the survival of restaurants by utilizing a comprehensive dataset from Yelp. The restaurant sector plays a significant role in the American economy, with the National Restaurant Association reporting that it contributed over $833 billion to the economy and provided employment to one out of every ten workers in 2018. The industry is also notorious for its high rate of turnover; Parsa et al. (2005) noted that the turnover rate for restaurants in their first year can reach 26%. Despite this, there is a lack of extensive empirical studies on the survival of restaurants. Our research particularly examines customer-generated reviews and ratings, along with various business attributes, to forecast a restaurant's longevity. We define business survival as the continued operation of a business, as opposed to its failure. The primary aim is to build machine learning models to predict the failure of real-world businesses and to provide actionable insights for small business owners, especially those in the restaurant sector. Our data is sourced from Yelp, the largest platform for consumer reviews, where customers express their opinions on restaurants. The dataset encompasses 150,346 restaurants across 1,416 cities in 27 U.S. states.

In the remainder of this report, we first discuss related works in Section 2, and describe our data and the definition of our variables in Section 3. We summarize the methods we used to do prediction in Section 4 and discuss our results in Section 5. Section 5 contains the conclusion and possible direction for future work.

**2 Related Works**

**2.1 Overview of business survival/failure prediction**

Business failure occurs when a company is forced to halt operations or shut down due to business distress, a condition characterized by dwindling financial or human resources, leading some to eventual collapse (Amankwah-Amoah, 2016; Amankwah-Amoah & Wang, 2019). Various factors contribute to business failure, including company size, type, industry specifics, entrepreneurial traits, and financial struggles (Liahmad et al., 2021; Mayr et al., 2021; Ucbasaran et al., 2013). The likelihood of a business's success or failure also heavily depends on its industry context (Agarwal & Gort, 1996; Audretsch & Mahmood, 1995; Opstad & Valenta, 2022). For instance, the restaurant and airline sectors typically see shorter survival times and lower survival rates compared to tech or pharmaceutical industries (Hensler et al., 1997). High entry rates and geographical concentration, as seen in the restaurant industry, correlate with lower survival rates (English et al., 1996), with competition identified as a key factor in restaurant failures (Wu et al., 2021), leaving those without distinct competitive edges more prone to failure.

Previous research in finance has extensively explored business failure or bankruptcy prediction, developing various statistical models for this purpose. Traditional business failure prediction (BFP) models, like discriminant and logit analysis, use financial ratios and covariates from financial reports to estimate bankruptcy risk (Bunyaminu et al., 2019; Gepp & Kumar, 2008). Such statistical analyses aid organizations in understanding their position and making informed decisions (Lee, 2014; Wieprow & Gawlik, 2021; Yang et al., 2011). However, these models often classify business status simply as active or liquidating, focusing more on financial management while offering limited operational strategy insights. Shifting from bankruptcy forecasting, survival analysis provides an alternative BFP technique, analyzing the timeline to business failure and suggesting a company's lifespan. For example, Naumzik et al. (2022) utilized review ratings as indicators of customer satisfaction to predict restaurant failure likelihood and risk state, finding improved prediction accuracy with survival analysis over traditional methods. Despite its advantages, survival analysis requires strict data assumption adherence for reliable predictions (Moncada-Torres et al., 2021), highlighting the need for more adaptable methods for handling extensive online data.

**2.2 Survival/failure in the restaurant industry**

Lee (1987) explored various factors that contribute to the success and failure in the restaurant industry, highlighting the importance of aspects like food quality, consistency, franchising, adaptability, and marketing strategies, including advertising and management practices. The role of effective promotion and marketing strategies has also been emphasized as crucial for the prosperity of restaurants (English et al., 1996). Nizam (2017) pointed out that robust marketing approaches are particularly essential for small, independent restaurants that face a higher risk of failure. Furthermore, engaging with consumers and addressing their preferences and needs is vital for the growth of restaurant businesses (Wang & Kim, 2021), as these businesses are heavily dependent on customer experiences and perceptions. Cant and Erdis (2012) conducted a survey-based exploratory study to pinpoint the factors linked to restaurant success from the consumers' perspective. Their findings highlighted cleanliness, value for money, service quality, and food quality as key elements in boosting customer satisfaction, fostering customer relationships, and thereby improving profitability and longevity of restaurant businesses. Parsa et al. (2011) applied the Kaplan–Meier method and Cox's hazard regression models in a survival analysis of restaurants, revealing the significant impact of factors such as location, affiliation, and size on restaurant failure and survival rates. Additionally, restaurants with larger sizes and more complex operations were shown to have better survival prospects (Parsa et al., 2011).

Expanding on prior studies, our objective is to utilize customer-generated reviews and ratings, combined with various business attributes, to develop a machine learning model that forecasts restaurant survival. This approach offers small businesses a flexible strategy for making decisions based on data analysis.

**3 Methodology**

**3.1 Data Description**

We collect the dataset from Yelp.com[[1]](#footnote-1). The Yelp dataset is a subset of Yelp’s businesses, reviews, and user data for academic purpose. This Yelp dataset is composed of six JSON files, and for our project, we aim to focus on three specific files: business data, tip data, and review data. Our primary dataset is the business information file, where each entry represents a business entity, including 14 attributes such as name, location, and opening hours. The business star rating serves as the dependent variable in our analysis, with other business attributes utilized as feature variables for model building. In addition to fundamental business attributes, we also intend to incorporate user review information. Initially, we'll utilize tip data, containing succinct suggestions from users about a business, offering quick insights due to their brevity. Furthermore, we plan to integrate user review data, including full review text, user\_id, and business\_id for each review. The details of each variable are listed in Table 1. The Yelp dataset, comprising 6.9 million reviews, 908,915 tips, 1.9 million users, and 150,346 businesses, is large, robust, and serves as a suitable sample for utilizing Machine Learning in predicting restaurant survival.

We cleaned and preprocessed the data. Table 1 describes the variables used in our analysis. The data was processed in the following manner: 1) Different datasets were merged using unique identifiers. 2) Binary variables were converted to integers. 3) Inconsistent data entries were standardized. For instance, prefixes and leading/trailing single quotes were removed from the "WiFi" categorical variable, which was then categorized into four levels: "nan," "no," "free," and "paid." 4) Missing values were assessed. 5) We randomly sampled 5% of the data due to its extensive size. Our dependent variable is free from missing values, whereas some restaurant characteristic features do have missing values. We interpreted these missing values as potentially informative for predictions, as the absence of certain information on Yelp could influence a customer's perception or decision regarding a restaurant. Therefore, we decided to retain all missing values.

Table 1: Variables Description

|  |  |
| --- | --- |
| Variable | Description |
| *is\_open* | integer, 0 or 1 for closed or open, respectively |
| *stars* | integer, star rating |
| *review\_count* | Integer, total review received |
| *ByAppointmentOnly* | Categorical, whether it is by appointment only |
| *BusinessAcceptsCreditCards* | Categorical, whether accept credit card, |
| *BikeParking* | Categorical, whether bike parking is available |
| *WiFi* | Categorical, whether it provide free, paid or no wifi |
| *HasTV* | Categorical, whether it has TV or not |

Table 2: Summary Statistics

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Numeric Variable | | | | |
| **Variable** | **Count** | **Min** | **Max** | **Mean** |
| is\_open | 150,346 | 0 | 1 | 0.79 |
| stars | 150,346 | 1 | 5 | 3.59 |
| review\_count | 150,346 | 5 | 7,568 | 44.87 |
| Categorical Variable | | | | |
| **Variable** | **Frequency** | **Percentage** |  |  |
| ByAppointmentOnly |  |  |  |  |
| True | 15,609 | 0.10 |  |  |
| False | 26,690 | 0.18 |  |  |
| Nan | 108,047 | 0.72 |  |  |
| BusinessAcceptsCreditCards |  |  |  |  |
| True | 113,667 | 0.76 |  |  |
| False | 6,098 | 0.04 |  |  |
| Nan | 30,581 | 0.20 |  |  |
| BikeParking |  |  |  |  |
| True | 55,040 | 0.37 |  |  |
| False | 17,518 | 0.12 |  |  |
| Nan | 77,788 | 0.51 |  |  |
| HasTV |  |  |  |  |
| True | 34,154 | 0.23 |  |  |
| False | 10,911 | 0.07 |  |  |
| Nan | 105,281 | 0.70 |  |  |
| WiFi |  |  |  |  |
| Free | 34,414 | 0.23 |  |  |
| Paid | 619 | 0.01 |  |  |
| No | 21,881 | 0.14 |  |  |
| Nan | 93,432 | 0.62 |  |  |

**3.2 Machine Learning Models**

**3.2.1 Model Specification**

**Logistic Model.** The logistic model, also known as logistic regression, is specifically designed for binary classification tasks. It models the probability that a given input belongs to a particular category (e.g., a restaurant will survive or fail) by using the logistic function to ensure that the output lies between 0 and 1. This model specified as follows:

|  |  |
| --- | --- |
|  | (1) |

This model is particularly suitable for binary classification tasks, such as predicting whether a restaurant will survive or not. It provides insights into how the presence or absence of certain features (e.g., high customer ratings, location) influences the likelihood of restaurant survival. Logistic regression is highly interpretable, like linear regression, but it is more appropriate for binary outcomes.

**Probit Model**. The probit model is similar to logistic regression but uses the cumulative distribution function of the normal distribution to link the linear predictors to the probability of the binary outcome. This makes the probit model particularly useful when the underlying latent variable is assumed to be normally distributed. The Model specified as follows:

|  |  |
| --- | --- |
|  | (2) |

Probit models can be used interchangeably with logistic models for binary classification problems. The choice between a logistic and a probit model often comes down to the distributional assumptions about the data and the error terms. The probit model's reliance on the normal cumulative distribution function makes it a good fit for scenarios where the latent variable (e.g., the propensity of a restaurant to succeed or fail) follows a normal distribution. It can sometimes provide a better fit to certain types of data compared to logistic regression.

**Support Vector Machine (SVM).** SVM is a supervised machine learning algorithm used for classification tasks. SVM works by finding the optimal hyperplane that best separates different classes in the feature space while maximizing the margin between the classes. It aims to find the decision boundary that not only separates the classes but also generalizes well to unseen data. SVM can handle both linear and non-linear data using kernel functions, which map the input space into a higher-dimensional space where it becomes easier to find a separating hyperplane. SVM is widely used in various fields such as image recognition, text classification, and bioinformatics due to its effectiveness and versatility.

**Linear Discriminant Analysis/Quadratic Discriminant Analysis.** Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA) are both classification models and dimensionality reduction techniques commonly employed in machine learning. LDA operates under the assumption that the probability distribution of input features is Gaussian and that each class shares the same covariance matrix. This assumption enables LDA to efficiently separate classes by finding linear combinations of features that maximize class separability. On the other hand, QDA relaxes the constraint of shared covariance matrices, allowing each class to have its own covariance matrix. This flexibility in QDA enables it to capture more complex decision boundaries, making it suitable for situations where the classes exhibit different variances or covariances. While LDA is computationally simpler and more interpretable, QDA provides a more flexible modeling approach, making the choice between the two dependent on the specific characteristics of the dataset and the desired trade-offs between complexity and performance.

**Decision Tree.** A Decision Tree is a hierarchical model commonly used for classification tasks. In this tree-structured model, internal nodes represent the features or attributes of a dataset, such as age, income, or gender, while branches emanating from these nodes depict decision rules based on the feature values. As the algorithm progresses down the tree, it makes decisions at each node based on the feature values until reaching a leaf node, which corresponds to the final prediction or outcome.

**Non-linearity**. We also consider adding non-linearity into the model’s prediction. First, this approach allows the model to capture more complex relationships between the features and the response variable than would be possible with linear terms alone. Second, including quadratic terms can help in modeling the interaction effects between features. Third, quadratic terms can improve the fit of the model to the training data, reducing the bias in predictions. This can be particularly useful in cases where a purely linear model underfits the data, failing to capture essential patterns. Last, adding quadratic terms increases the flexibility of the model, allowing it to conform more closely to the underlying data structure.

**3.2.2 Model Evaluation**

We use accuracy to access model performance. A higher accuracy indicates a model that more accurately predicts the outcomes.

We first train the model by using training set and then test the model by using testing set. However, this method might not fully represent the model's ability to generalize across different subsets of the data because the assessment is based on a single split of the data. Therefore, we further evaluate the model performance by using 5-fold cross validation. In 5-fold CV, the training dataset is divided into five equal-sized subsets or 'folds'. The model is then trained and evaluated five times, each time using a different fold as the testing set and the remaining four folds as the training set. The final accuracy is calculated as the average of the accuracy from each of the five iterations.

**4 Results and Model Performance Evaluation**

Table 3 presents the testing errors for each model. The validation process indicates that the Logistic Model with Quadratic Terms outperforms the others with the highest accuracy score. In an evaluation using a single test set, incorporating a quadratic term resulted in diminished model performance. However, in a 5-fold cross-validation framework, the addition of a quadratic term enhanced the model's performance. The difference in performance when adding a quadratic term between testing with a single test set and using 5-fold cross-validation can be attributed to overfitting. With a single test set, overfitting may occur, leading to decreased performance due to the model capturing noise in the data. In contrast, 5-fold cross-validation mitigates overfitting by evaluating the model on multiple subsets of the data, providing a more reliable estimate of its generalization performance. If the quadratic term improves performance during cross-validation, it suggests that it captures meaningful patterns in the data that generalize well across different subsets. Despite the validation leading to the same conclusion regarding model selection, the testing accuracy derived from 5-fold cross-validation are somewhat lower than those obtained from a single test set. This suggests that 5-fold cross-validation might offer a more accurate evaluation of a model's performance on unseen data.

Table 3 : Model Performance Comparison

|  |  |
| --- | --- |
| Model | Testing Accuracy |
| *Testing on Single Testing Set* |  |
| Logistic Model | **0.8087** |
| Logistic Model with Quadratic Terms | 0.8057 |
| Probit Model | **0.8087** |
| Probit Model with Quadratic Terms | 0.8050 |
| SVM with Linear Kernel | 0.7954 |
| SVM with Radial Basis Kernel | 0.8083 |
| LDA | 0.8063 |
| QDA | 0.7438 |
| Decision Tree | 0.7954 |
| *5-fold Cross Validation* |  |
| Logistic Model | 0.8035 |
| Logistic Model with Quadratic Terms | **0.8037** |
| Probit Model | 0.8025 |
| Probit Model with Quadratic Terms | 0.8029 |
| SVM with Linear Kernel | 0.7954 |
| SVM with Radial Basis Kernel | 0.8007 |
| LDA | 0.8008 |
| QDA | 0.7390 |
| Decision Tree | 0.7954 |

**5 Conclusion and Future Work**

In conclusion, based on the accuracy evaluated through 5-fold cross-validation, it is evident that the logistic model with a quadratic term outperforms its counterpart without. This superiority suggests the presence of non-linear relationships between the features and the response variable. By incorporating the quadratic term, the model can better capture these intricate patterns in the data, leading to improved accuracy. This finding underscores the importance of considering non-linear relationships in predictive modeling, as they can significantly impact model performance and enhance predictive capabilities. Therefore, leveraging models that can accommodate non-linearities, such as logistic regression with quadratic terms, is essential for accurately capturing the complexities inherent in real-world datasets.

**Incorporation of Additional Machine Learning Models**. To broaden our analysis, we will explore a variety of advanced machine learning models, particularly those based on decision trees, such as Random Forest and Gradient Boosting Machines (GBMs). These models are known for their ability to handle non-linear relationships and complex interactions between features, making them potentially more effective in predicting restaurant survival. By comparing their performance against the baseline models, we can identify the most suitable approach for our dataset.

**Expansion of Feature Set**. Recognizing the importance of comprehensive feature representation, we plan to incorporate additional features into our models. This may include more detailed customer sentiment analysis from reviews, geographical factors, competitive landscape analysis, and temporal features such as seasonality. Enriching the feature set will help capture a wider array of factors that might influence restaurant survival, potentially improving model accuracy.

**Analysis of Feature Importance**. To derive practical insights from our models, we will conduct an analysis of feature importance. This involves determining which features contribute most significantly to the model's predictions, which can offer valuable insights for restaurant owners and managers. Understanding key predictors of survival can inform strategic decisions, such as focusing on customer service improvement, adjusting marketing strategies, or optimizing menu offerings.

**6 Reference**

Anagnostopoulou, S. C., Buhalis, D., Kountouri, I. L., Manousakis, E. G., & Tsekrekos, A. E. (2019). The impact of online reputation on hotel profitability. International Journal of Contemporary Hospitality Management, 32(1), 20-39.

Altman, E. I. (1967). The prediction of corporate bankruptcy: A discriminant analysis. University of California, Los Angeles.

Amankwah-Amoah, J. (2016). An integrative process model of organisational failure. Journal of Business Research, 69(9), 3388-3397.

Audretsch, D. B., & Mahmood, T. (1995). New firm survival: new results using a hazard function. The review of economics and statistics, 97-103.

Amankwah-Amoah, J., & Wang, X. (2019). Business failures around the world: Emerging trends and new research agenda. Journal of Business Research, 98, 367-369.

Agarwal, R., & Gort, M. (1996). The evolution of markets and entry, exit and survival of firms. The review of Economics and Statistics, 489-498.

Beaver, W. H. (1966). Financial ratios as predictors of failure. Journal of accounting research, 71-111.

Bunyaminu, A., Mohammed, I., & Issah, M. (2019). Business failure prediction: A tri-dimensional approach. The Journal of Applied Business and Economics, 21(2), 80-100.

Cant, M. C., & Erdis, C. (2012). Incorporating customer service expectations in the restaurant industry: The guide to survival. Journal of Applied Business Research (JABR), 28(5), 931-942.

Duan, W., Gu, B., & Whinston, A. B. (2008). Do online reviews matter?—An empirical investigation of panel data. Decision support systems, 45(4), 1007-1016.

English, W. (1996). Restaurant attrition: a longitudinal analysis of restaurant failures. International Journal of Contemporary Hospitality Management, 8(2), 17-20.

Gepp, A., & Kumar, K. (2008). The role of survival analysis in financial distress prediction. International research journal of finance and economics, 16(16), 13-34.

Hensler, D. A., Rutherford, R. C., & Springer, T. M. (1997). The survival of initial public offerings in the aftermarket. Journal of Financial Research, 20(1), 93-110.

Liahmad, K. R., Utami, Y. P., & Sitompul, S. (2021). Financial factors and non-financial to financial distress insurance companies that listed in Indonesia Stock Exchange. Budapest International Research and Critics Institute (BIRCI-Journal): Humanities and Social Sciences, 4(1), 1305-1312.

Lee, M. C. (2014). Business bankruptcy prediction based on survival analysis approach. AIRCC's International Journal of Computer Science and Information Technology, 103-119.

Lee, D. R. (1987). Why some succeed where others fail. Cornell Hotel and Restaurant Administration Quarterly, 28(3), 32-37.

Mayr, S., Mitter, C., Kücher, A., & Duller, C. (2021). Entrepreneur characteristics and differences in reasons for business failure: evidence from bankrupt Austrian SMEs. Journal of Small Business & Entrepreneurship, 33(5), 539-558.

Moncada-Torres, A., van Maaren, M. C., Hendriks, M. P., Siesling, S., & Geleijnse, G. (2021). Explainable machine learning can outperform Cox regression predictions and provide insights in breast cancer survival. Scientific reports, 11(1), 6968.

Nieto, J., Hernández-Maestro, R. M., & Muñoz-Gallego, P. A. (2014). Marketing decisions, customer reviews, and business performance: The use of the Toprural website by Spanish rural lodging establishments. Tourism management, 45, 115-123.

Nizam, H. (2017). Survival strategies for small independent full-service restaurants. Walden University.

Naumzik, C., Feuerriegel, S., & Weinmann, M. (2022). I will survive: Predicting business failures from customer ratings. Marketing Science, 41(1), 188-207.

Opstad, L., & Valenta, R. (2022). The Long-Run Equilibrium of Industry Population and Bottom Lines: Analysing the Near-Perfect Restaurant Industry. Journal of Applied Business and Economics, 24(3).

Parsa HG, Gregory A, Terry MD (2010) Why do restaurants fail? Part III: An analysis of macro and micro factors. Emerging Aspects Redefining Tourism Hospitality 1(1):16–25.

Parsa, H. G., Self, J., Sydnor-Busso, S., & Yoon, H. J. (2011). Why restaurants fail? Part II-The impact of affiliation, location, and size on restaurant failures: Results from a survival analysis. Journal of Foodservice Business Research, 14(4), 360-379.

Phillips, P., Barnes, S., Zigan, K., & Schegg, R. (2017). Understanding the impact of online reviews on hotel performance: an empirical analysis. Journal of Travel Research, 56(2), 235-249.

Sarkar, S., & Sriram, R. S. (2001). Bayesian models for early warning of bank failures. Management Science, 47(11), 1457-1475.

Torres, E. N., Singh, D., & Robertson-Ring, A. (2015). Consumer reviews and the creation of booking transaction value: Lessons from the hotel industry. International Journal of Hospitality Management, 50, 77-83.

Ucbasaran, D., Shepherd, D. A., Lockett, A., & Lyon, S. J. (2013). Life after business failure: The process and consequences of business failure for entrepreneurs. Journal of management, 39(1), 163-202.

Viglia, G., Minazzi, R., & Buhalis, D. (2016). The influence of e-word-of-mouth on hotel occupancy rate. International Journal of Contemporary Hospitality Management, 28(9), 2035-2051.

Wang, Y., & Kim, J. (2021). Interconnectedness between online review valence, brand, and restaurant performance. Journal of Hospitality and Tourism Management, 48, 138-145.

Wu, M., Pei, T., Wang, W., Guo, S., Song, C., Chen, J., & Zhou, C. (2021). Roles of locational factors in the rise and fall of restaurants: A case study of Beijing with POI data. Cities, 113, 103185.

Wieprow, J., & Gawlik, A. (2021). The use of discriminant analysis to assess the risk of bankruptcy of enterprises in crisis conditions using the example of the tourism sector in Poland. Risks, 9(4), 78.

Yang, Z., You, W., & Ji, G. (2011). Using partial least squares and support vector machines for bankruptcy prediction. Expert Systems with Applications, 38(7), 8336-8342.

|  |  |
| --- | --- |
| Model | Accuracy |
| Logistic Model | 0.8035 |
| Logistic Model with Quadratic Terms | 0.8037 |
| Probit Model | 0.8025 |
| Probit Model with Quadratic Terms | 0.8029 |
| SVM with Linear Kernel | 0.7954 |
| SVM with Radial Basis Kernel | 0.8007 |
| LDA Model | 0.8008 |
| QDA Model | 0.7391 |
| Decision Tree Model | 0.7954 |

1. Yelp Open Dataset. <https://www.yelp.com/dataset> [↑](#footnote-ref-1)